

1.1 Title and Investigators

- Title: Machine Learning for Earth Science Modeling
- Principal Investigator: Mark Schwabacher, NASA Ames, Code IC
- In collaboration with: Rama Nemani and Dave Opitz, University of Montana

1.2. Goal & Technical Objectives

- Improve fit of forest ecology models to satellite data and ground-based data
- Improve ability to predict forest fires
- Improve ability to plan data-processing operations
- Discover scientifically interesting patterns in the data
- Discover errors in the data

1.3. Technical Problem Statement

- We need to find and/or develop the right machine learning and data mining algorithms to automatically learn or discover from the data the types of things that will be useful to Earth science research.

1.4. Technical Approach

- Look for ways in which machine learning and data mining technology can be applied to each of the research objectives using the data sets that are available at the University of Montana, and then experiment with applying these technologies to these data sets. Develop new algorithms as necessary.

1.5. Data and NASA Relevance

- The project is using data from the Numerical Terradynamic Simulation Group in the Forestry Department at the University of Montana
- The project seeks to improve forest ecology models, in support of NASA's goal of improving our understanding of the Earth.

1.6. Accomplishments & Preliminary Findings

- Meeting held at University of Montana in Missoula during which we identified five ways to apply machine learning to forest ecology, and decided to pursue the data mining approach first based on data availability
- Dave Opitz and his students at UMT have started to build an interactive software system to do data mining on the ecological data
- Using this system, an error in the data was discovered: freezing temperatures in the summer in the Middle East

1.7. Technical Significance of Progress and Expected Impact on NASA

- By identifying errors in the ecological data, the errors can be corrected, resulting in increases in the quality of the data which can result in improved understanding of the Earth

1.8. Linkable URL's which describe your team and/or your research

- Dr. Mark Schwabacher's home page:
<http://www0.arc.nasa.gov/~schwabac/>
- Web page of the Numerical Terradynamic Simulation Group in the Forestry Department at the University of Montana:
<http://www.forestry.umt.edu/ntsg/>

1.9. Facilities Used and Personnel Assigned to Projects (students, postdoc, contractors, etc.)

- Dr. Mark Schwabacher, NASA ARC, Code IC

1.10 References

- M. Schwabacher and P. Langley.
Discovering Communicable Scientific
Knowledge from Spatio-Temporal Data.
International Conference on Machine
Learning. July 2001.